**CAPSTONE PROJECT PROPOSAL**

**GROUP 1**

SIDDHARTHA PATRA – 0787887

USAMA NORAT – 0792225

AMIR DAHYA - 0791252

**INTRODUCTION**

Recruiting the right candidates for a job is a crucial task for any company. However, the process of screening resumes can be time-consuming and error-prone. To automate this process, we propose to develop a Resume Parser using the NLP library ‘Spacy’, which can help to extract key information (or entities) from resumes, such as education, work experience, and skills.

**PROJECT OBJECTIVE**

The objective of the proposed project is to develop a Resume Parser that can quickly extract key information from resumes and match them with job requirements. We will use Spacy, which is a powerful NLP library in Python, to analyze resumes and extract relevant information. This information will be displayed in an easy-to-use web-based application that recruiters can use to match candidates with job requirements.

**PROJECT DESCRIPTION**

**PROBLEM STATEMENT:**

The problem we are addressing is the time-consuming and tedious task of manually parsing resumes for relevant information. This task is often performed by human resources departments or recruiters during the hiring process. The issue becomes particularly challenging when dealing with a large number of resumes.

**PROJECT SIGNIFICANCE:**

The problem we are dealing in this project is important because from a business point-of-view, it can significantly slow down the hiring process and leads to missed opportunities. Automating the resume parsing process can save time and increase efficiency, allowing HR departments and recruiters to focus on other important tasks such as interviewing and hiring the best candidates.

**PROPOSED SOLUTION:**

Our solution is to build a resume parser using the Spacy NLP python library. This parser will be able to extract key entities such as work experience, education, skills etc. from resumes. The parser will then be deployed as a web-based application on a cloud-based platform such as GCP which can allow to scale up the application to easily upload a bulk of resumes and parse resumes through a user-friendly interface.

**STEPS INVOLVED IN THE PROJECT**:

1. Developing a Python-based NLP model that can extract key information (or entities) from resumes using Spacy
2. Training the model using a large dataset of resumes
3. Building a web-based application using a web framework (e.g., Flask, Django) that can accept resumes in various formats (e.g., PDF, Word, etc.) and parse the information using the trained model
4. Integrating the application with a cloud-based platform (such as Google Cloud Platform) for deployment
5. Providing an easy-to-use interface for recruiters to access the parsed information and match it with the job requirements

**KNOWLEDGE/SKILLS/TECH NEEDED:**

* Knowledge of natural language processing (NLP) concepts such Tokenization, Lemmatization, Parts-of-Speech Tagging etc.
* Experience with the Spacy NLP library
* Knowledge of Python programming
* Experience with web-based application development and deployment on cloud-based platforms

**COMPARISON TO EXISTING SOLUTIONS:**

There are a number of existing resume parsing solutions on the market, but our solution stands out in its use of the Spacy NLP library. Spacy is a widely used library in the NLP community and is known for its speed and accuracy. Additionally, our solution will be deployed as a web-based application, making it end-to-end data science project.

**EVALUATION METRICS:**

* Accuracy of resume entity extraction
* Time taken for the information extraction process
* Number of resumes parsed

**MEASUREMENT OF SUCCESS:**

The success of the solution will be measured by how well it can automate the resume review process and how much time and resources it can save for recruiters.

**EVALUATION AND FEASIBILITY:**

The solution will be evaluated by testing it on a sample set of college student resumes and receiving feedback from our faculty. The feasibility of the solution is high as it is based on proven technologies and techniques.

**ESTIMATED IMPACT:**

The impact of this solution is expected to be significant as it will significantly speed up the hiring process and increase efficiency for HR departments and recruiters. Additionally, it will make the job application process more efficient and streamlined for job seekers.

**DATA SOURCE:**

For the purpose of training, 220 resumes were downloaded from an online jobs platform. These documents were uploaded to [Dataturks](https://docs.dataturks.com/) online annotation tool and manually annotated. The tool automatically parses the documents and allows for us to create annotations of important entities we are interested in and generates JSON formatted training data with each line containing the text corpus along with the annotations. The above dataset consisting of 220 annotated resumes can be found [here](https://dataturks.com/projects/abhishek.narayanan/Entity%20Recognition%20in%20Resumes).

**CONCLUSION**

In conclusion, our proposed project aims to develop a Resume Parser that can significantly reduce the time and effort required to screen resumes, increase the efficiency of the recruitment process, and help companies find the right candidate for the job. By using Spacy NLP library and deploying the application on a cloud-based platform, the solution can be easily integrated with existing recruitment software and scaled up to meet the needs of any organization. We believe that our solution can bring a significant impact to the recruitment process and make the process more efficient for companies.